

AABC: approximate approximate Bayesian computation when simulating a large number of data sets is computationally infeasible

Erkan O. Buzbas

Department of Biology

Stanford University, Stanford, CA 94305-5020 USA

and

Department of Statistical Science

University of Idaho, Moscow, ID 84844-1104 USA

email: erkanb@uidaho.edu

and

Noah A. Rosenberg

Department of Biology

Stanford University, Stanford, CA 94305-5020 USA

email: noahr@stanford.edu

January 29, 2013

Abstract

Approximate Bayesian computation (ABC) methods perform inference on model-specific parameters of mechanistically motivated parametric statistical models when evaluating likelihoods is difficult. Central to the success of ABC methods is computationally inexpensive simulation of data sets from the parametric model of interest. However, when simulating data sets from a model is so computationally expensive that the posterior distribution of parameters cannot be adequately sampled by ABC, inference is not straightforward. We present “approximate approximate Bayesian computation” (AABC), a class of methods that extends simulation-based inference by ABC to models in which simulating data is expensive. In AABC, we first simulate a *limited number* of data sets that is computationally feasible to simulate from the parametric model. We use these data sets as fixed background information to inform a non-mechanistic statistical model that approximates the correct parametric model and enables efficient simulation of a large number of data sets by Bayesian resampling methods. We show that under mild assumptions, the posterior distribution obtained by AABC converges to the posterior distribution obtained by ABC, as the number of data sets simulated from the parametric model and the sample size of the observed data set increase simultaneously. We illustrate the performance of AABC on a population-genetic model of natural selection, as well as on a model of the admixture history of hybrid populations.

KEYWORDS: Approximate Bayesian computation, likelihood-free methods, nonparametrics, posterior distribution

1 Introduction

Stochastic processes motivated by mechanistic considerations enable investigators to capture salient phenomena in modeling natural systems. Statistical models resulting from these stochastic processes are often parametric, and estimating model-specific parameters—which often have a natural interpretation—is a major aim of data analysis. Contemporary mechanistic models tend to involve complex stochastic processes, however, and parametric statistical models resulting from these processes lead to computationally intractable likelihood functions. When likelihood functions are computationally intractable, likelihood-based inference is a challenging problem that has received considerable attention in the literature (Robert and Casella, 2004; Liu, 2008).

When statistical models are known only at the level of the stochastic mechanism generating the data—such as in implicit statistical models (Diggle and Gratton, 1984)—explicit evaluation of likelihoods might be impossible. In these models, standard computational methods that require evaluation of likelihoods up to a proportionality constant (e.g., rejection methods) cannot be used to sample distributions of interest. However, data sets simulated from the model under a range of parameter values can be used to assess parameter likelihoods without explicit evaluation (Rubin, 1984). Approximate Bayesian computation (ABC) methods (Tavaré *et al.*, 1997; Beaumont *et al.*, 2002; Marjoram *et al.*, 2003) implement this idea in a Bayesian context to sample an *approximate* posterior distribution of the parameters. Intuitively, parameter values producing simulated data sets similar to the observed data set arise in approximate proportion to their likelihood, and hence, when weighted by prior probabilities, to their posterior probabilities.

1.1 The ABC literature

ABC methods have been based on rejection algorithms (Tavaré *et al.*, 1997; Beaumont *et al.*, 2002; Blum and François, 2010), Markov chain Monte Carlo (Beaumont, 2003; Marjoram *et al.*, 2003; Bortot *et al.*, 2007; Wegmann *et al.*, 2009), and sequential Monte Carlo (Sisson *et al.*, 2007, 2009; Beaumont *et al.*, 2009; Toni *et al.*, 2009). Model selection using ABC (Pritchard *et al.*, 1999; Fagundes *et al.*, 2007; Grelaud *et al.*, 2009; Blum and Jakobsson, 2010; Robert *et al.*, 2011), the choice of summary statistics when the likelihood is based on summary statistics instead of the full data (Joyce and Marjo-

ram, 2008; Wegmann *et al.*, 2009; Nunes and Balding, 2010; Fearnhead and Prangle, 2012), and the equivalence of posterior distributions targeted in different ABC methods (Wilkinson, 2008; Sisson *et al.*, 2010) have also been investigated.

ABC methods have had a considerable effect on model-based inference in disciplines that rely on genetic data, particularly data shaped by diverse evolutionary, demographic, and environmental forces. Example applications have included problems in the demographic history of populations (Pritchard *et al.*, 1999; François *et al.*, 2008; Verdu *et al.*, 2009; Blum and Jakobsson, 2010) and species (Estoup *et al.*, 2004; Plagnol and Tavaré, 2004; Becquet and Przeworski, 2007; Fagundes *et al.*, 2007; Wilkinson *et al.*, 2010), as well as problems in the evolution of cancer cell lineages (Tavaré, 2005; Siegmund *et al.*, 2008) and the evolution of protein networks (Ratmann *et al.*, 2009). Other applications outside of genetics have included inference on the physics of stereological extremes (Bortot *et al.*, 2007), the ecology of tropical forests (Jabot and Chave, 2009), dynamical systems in biology (Toni *et al.*, 2009), and small-world network disease models (Walker *et al.*, 2010). ABC methods have been reviewed by Marjoram and Tavaré (2006), Cornuet *et al.* (2008), Beaumont *et al.* (2009), Beaumont (2010), Csilléry *et al.* (2010), and Marin *et al.* (2011).

1.2 A limitation of ABC methods

An informal categorization of the information available about the likelihood function is helpful to illustrate the class of models in which ABC methods are most useful. First, exact inference on the posterior distribution of the parameters is possible only if the likelihood function is analytically available. Second, if the likelihood function is not analytically available but can be evaluated up to a constant given a parameter value, then standard computational methods such as rejection algorithms can sample the posterior distribution. In this case, inference is exact up to a Monte Carlo error due to sampling from the posterior. Third, if the likelihood function cannot be evaluated, but data sets can feasibly be simulated from the model, then ABC methods sample the posterior distribution using approximations on the *data space* in addition to a Monte Carlo error due to sampling.

Although ABC methods sample the posterior distribution of parameters without evaluating the likelihood function, they are computationally intensive. Adequately sampling a posterior distribution of a parameter by ABC

requires many random realizations from the prior distribution of the parameter and the sampling distribution of the data. Simulating from the prior is straightforward, but the computational cost of simulating a data set from the mechanistic model increases quickly with the complexity and number of stochastic processes involved. Henceforth, we refer to statistical models in which not only evaluating the likelihoods is difficult but also simulating a large number of data sets is computationally infeasible as *limited-generative* models. When a model is limited-generative and only a small number of data sets can be simulated from the model, likelihoods cannot be assessed using ABC and hence, the posterior distribution of parameters cannot be adequately sampled.

1.3 Our contribution

In this article, we introduce *approximate* approximate Bayesian computation (AABC), a class of methods that perform inference on model-specific parameters of limited-generative models when standard ABC methods are computationally infeasible to apply. In AABC, the idea of assessing the likelihoods approximately using simulated data sets is taken one step further than in ABC. AABC methods make approximations on the *parameter space* and the *model space* in addition to standard ABC approximations on the data space. In conjunction with Bayesian resampling methods, these approximations help us overcome the computational intractability associated with simulating data from a limited-generative model (Figure 1).

Our key innovation is to condition on a limited number of data sets that can be feasibly simulated from the limited-generative model and to employ a non-mechanistic statistical model to simulate a large number of data sets. We set up the non-mechanistic model based on empirical distributions of the limited number of data sets simulated from the mechanistic model. Since the data values from the limited number of simulated data sets are used to construct new random data sets by resampling methods, it is computationally inexpensive to simulate a large number of data sets in AABC. The AABC approach allows a researcher to allocate a fixed computer time to simulating a limited number of data sets from the limited-generative model, thus making otherwise challenging likelihood-based inference attainable.

Intuitively, the information conditioned upon by the non-mechanistic model increases with the number of data sets simulated from the mechanistic model, and the expected accuracy of inference obtained by AABC

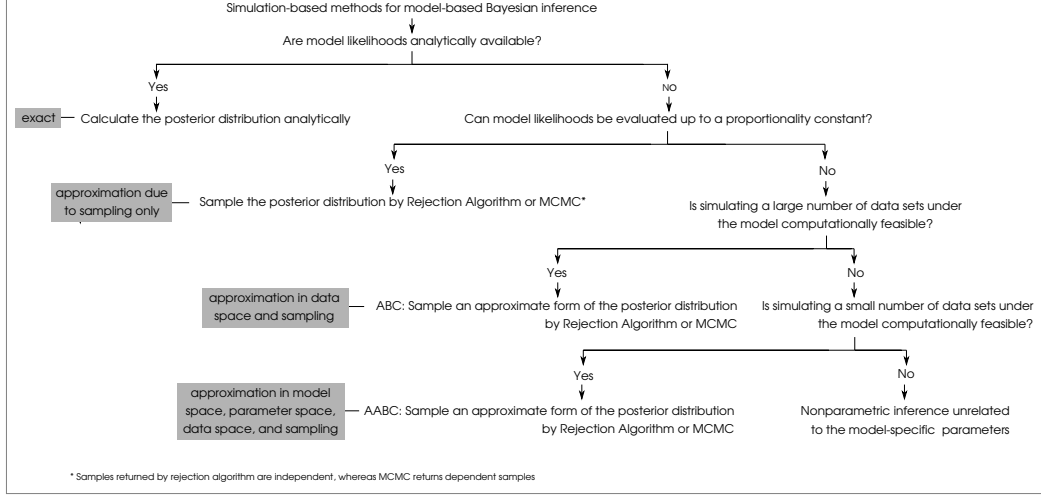


Figure 1: Applicability of simulation-based inference methods in relation to the information available about the likelihood function.

methods increases. We formalize this intuition by showing that the posterior distribution of parameters obtained by AABC converges to the corresponding posterior distribution obtained by standard ABC, as the sample size of the observed data set and the number of data sets simulated from the limited-generative model increase simultaneously.

AABC methods utilize the established machinery of ABC methods in sampling the posterior distribution of the parameters. Therefore, standard approximations on the data space involved in an ABC method—which facilitate the sampling of the posterior distribution—apply to AABC methods as well. We now briefly review these approximations in the context of ABC by rejection algorithms.

2 Review of ABC by rejection algorithms

To more formally set up the class of problems in which ABC methods are useful, we assume that a parametric model generates observations conditional on parameter $\theta \in \Theta \equiv \mathbf{R}^p$, $p \geq 1$. We let P_θ be the sampling distribution of a data set of n observations independent and identically distributed (IID) from this model. We denote a random data set by $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathcal{X}$,

where \mathcal{X} is the space in which the data set sits, and the observed data set by \mathbf{x}_o . In the genetics context, a data point x_i might be a vector denoting the allelic types of a genetic locus at genomic position i in a group of individuals; the data matrix \mathbf{x} might then contain genotypes from these individuals in a sample of n independent genetic loci.

Suppose that P_θ is available to the extent that the likelihood function $p(\mathbf{x}_o|\theta)$ can be evaluated up to a constant whose value does not depend on the parameters. Given a prior distribution $\pi(\theta)$ on parameter θ , the posterior distribution of θ given the observed data \mathbf{x}_o under the model P_θ is $\pi(\theta|\mathbf{x}_o, P_\theta)$. Then $\pi(\theta|\mathbf{x}_o, P_\theta)$ can be sampled by standard rejection sampling from $p(\mathbf{x}_o|\theta)\pi(\theta)$, a quantity that is proportional to $\pi(\theta|\mathbf{x}_o, P_\theta)$ by Bayes' Theorem. In principle, sampling $\pi(\theta|\mathbf{x}_o, P_\theta)$ without evaluating the likelihood function $p(\mathbf{x}_o|\theta)$ is possible, if simulating the data from the model P_θ is feasible. An early example due to Tavaré *et al.* (1997) samples $\pi(\theta|\mathbf{x}_o, P_\theta)$ by accepting a value θ_i simulated from the prior $\pi(\theta)$ only if the data set \mathbf{x}_i simulated from P_{θ_i} satisfies $\mathbf{x}_i = \mathbf{x}_o$. By standard rejection algorithm arguments, the θ_i sampled in this fashion are from the correct posterior distribution. However, the acceptance condition $\mathbf{x}_i = \mathbf{x}_o$ is rarely satisfied with high-dimensional data. A first approximation in ABC methods is dimension reduction by substituting the data set \mathbf{x} with a low-dimensional set of summary statistics \mathbf{s} . The observed data \mathbf{x}_o and the simulated data \mathbf{x}_i are substituted by \mathbf{s}_o and \mathbf{s}_i , calculated from their respective data sets. This is equivalent to substituting the likelihood function of the data $p(\mathbf{x}|\theta)$ with the likelihood function of the summary statistics $p(\mathbf{s}|\theta)$. Since ABC is most useful in statistical models that do not admit sufficient statistics, dimension reduction to summary statistics often entails information loss about the parameters. The choice of summary statistics minimizing this information loss is an active research area (Joyce and Marjoram, 2008; Wegmann *et al.*, 2009; Robert *et al.*, 2011; Aeschbacher *et al.*, 2012; Fearnhead and Prangle, 2012).

When the data are substituted with summary statistics, the acceptance condition $\mathbf{x}_i = \mathbf{x}_o$ is substituted by $\mathbf{s}_i = \mathbf{s}_o$, but exact equality may still be too stringent a condition to be satisfied with simulated data. A second approximation in ABC is to relax the exact acceptance condition with a tolerance acceptance condition. For example, Pritchard *et al.* (1999) used the Euclidean distance $\|\cdot\|$ and a small tuning parameter ϵ to accept a value θ_i from an approximate posterior distribution if the data set \mathbf{x}_i simulated

from P_{θ_i} produced \mathbf{s}_i satisfying

$$\|\mathbf{s}_i - \mathbf{s}_o\| = \left[\sum_{j=1}^k (s_{ij} - s_{oj})^2 \right]^{1/2} \leq \epsilon, \quad (1)$$

where \mathbf{s} is a k -dimensional statistic, and s_{ij} and s_{oj} are the j th components of \mathbf{s}_i and \mathbf{s}_o , respectively (see also Weiss and Von Haeseler (1998) for an application in a pure likelihood inference context). Distance metrics other than the Euclidean distance, such as the total variation distance (Tavaré *et al.*, 2002), have also been used.

Substituting the binary accept/reject step in the rejection sampling by weighting \mathbf{s}_i smoothly according to its distance from \mathbf{s}_o using a kernel density $K_\epsilon(\mathbf{s}_i, \mathbf{s}_o)$ with bandwidth ϵ leads to importance sampling (Wilkinson, 2008). The tolerance condition $\|\mathbf{s}_i - \mathbf{s}_o\| \leq \epsilon$ in the rejection algorithm of Pritchard *et al.* (1999) then corresponds to using a uniform kernel on an ϵ -ball around \mathbf{s}_o . Other approaches to kernel choice include Epanechnikov (Beaumont *et al.*, 2002) and Gaussian (Leuenberger and Wegmann, 2010) kernels.

When the data likelihood is substituted by the likelihood based on the summary statistics and a tolerance condition with a uniform kernel and the Euclidean distance is used, the posterior distribution sampled with ABC by rejection is

$$\pi_\epsilon(\theta|\mathbf{x}_o, P_\theta) = \frac{1}{C_{P_\theta}} \int_{\mathcal{X}} \mathbf{I}_{\{\|\mathbf{s} - \mathbf{s}_o\| < \epsilon\}} p(\mathbf{x}|\theta) \pi(\theta) d\mathbf{x}, \quad (2)$$

where \mathbf{I}_A is an indicator function that takes a value of 1 on set A and is zero otherwise, and $C_{P_\theta} = \int_{\Theta} \int_{\mathcal{X}} \mathbf{I}_{\{\|\mathbf{s} - \mathbf{s}_o\| < \epsilon\}} p(\mathbf{x}|\theta) \pi(\theta) d\mathbf{x} d\theta$ is the normalizing constant. A standard ABC algorithm that samples $\pi_\epsilon(\theta|\mathbf{x}_o, P_\theta)$ appears in Figure 2.

The choice of summary statistics, tolerance parameter ϵ , distance function, and kernel constitute approximations on the data space in ABC methods. We assume that these standard ABC approximations work reasonably well, and we focus on new modeling approximations on the parameter and model spaces introduced by AABC (Figure 3).

Algorithm 1: ABC by rejection algorithm

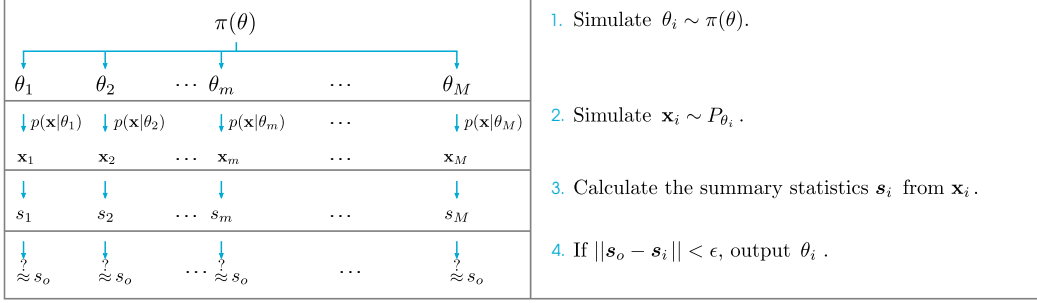


Figure 2: The ABC algorithm by rejection sampling. One iteration of the algorithm is shown on the right along with a schematic illustration of sampling from the posterior distribution of θ based on M proposed parameter values (left).

3 Approximate approximate Bayesian computation (AABC)

Algorithm 1 returns an adequate sample size from the posterior distribution of a parameter if it is iterated a large number of times, M . The set of realizations simulated from the joint distribution of the parameter and the data by steps 1 and 2 of Algorithm 1 is then $\{(\mathbf{x}_1, \theta_1), (\mathbf{x}_2, \theta_2), \dots, (\mathbf{x}_M, \theta_M)\}$. AABC methods seek inference on parameter θ when the model P_θ is limited-generative, and simulating M data sets under P_θ is therefore computationally infeasible. We thus assume that only a limited number m of data sets $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ can be obtained by step 2 of Algorithm 1 ($m \ll M$). We denote the set of realizations simulated from the joint distribution of the parameter and the data by $\mathcal{Z}_{n,m} = \{(\mathbf{x}_1, \theta_1), (\mathbf{x}_2, \theta_2), \dots, (\mathbf{x}_m, \theta_m)\}$, where each data set \mathbf{x}_i of n IID observations is simulated from the model P_{θ_i} .

In AABC, we substitute the joint sampling distribution P_θ of a data set of size n with the joint sampling distribution Q_θ , from which simulating data sets is computationally inexpensive. In replacing P_θ with Q_θ , we require that the posterior distribution $\pi(\theta|\mathbf{x}_o, Q_\theta)$ based on the likelihood implied by model Q_θ approximates the posterior distribution $\pi(\theta|\mathbf{x}_o, P_\theta)$ based on the likelihood implied by model P_θ . Further, we require that Q_θ can be used with a wide range of P_θ , in the sense that Q_θ is constructed without using

True quantity	Approximated by	Space involved	Source of error	Method employing approximation			
				Exact ABC with full data	Exact ABC with summary statistics	ABC	AABC
$\int_{\mathcal{X}} p(\mathbf{x} \theta)$	$\sum_{i=1}^M p(\mathbf{x}_i \theta)$	Data	Monte Carlo	yes	yes	yes	yes
$p(\mathbf{x} \theta)$	$p(\mathbf{s} \theta)$	Data	Dimension reduction	no	yes	yes	yes
$p(\mathbf{s}_o \theta)$	$p(\ \mathbf{s}_o - \mathbf{s}_i\ < \epsilon \theta)$	Summary statistics	Tolerance kernel Distance function	no	no	yes	yes
$p(\mathbf{s}_o \theta)$	$p(\mathbf{s}_o \tilde{\theta})$	Parameter	Tolerance Kernel Distance function	no	no	no	yes
$p(\mathbf{x} \theta)$	$\hat{p}(\mathbf{x} \theta)$	Model	Empirical distribution	no	no	no	yes

Figure 3: Approximations and errors involved in simulation-based ABC inference methods. Likelihood functions of the full data and the summary statistics are denoted respectively by $p(\mathbf{x}|\theta)$ and $p(\mathbf{s}|\theta)$. Exact ABC with full data involves only the Monte Carlo approximation due to sampling and thus is equivalent to a standard rejection algorithm. Summary statistics \mathbf{s} are assumed not to be sufficient so that dimension reduction from \mathbf{x}_o to \mathbf{s}_o results in an approximation.

the details of model P_θ .

3.1 Approximations on the parameter and model spaces due to replacing P_θ with Q_θ

Two approximations are involved in substituting P_θ with Q_θ . First, $\mathcal{Z}_{n,m}$ includes only m parameter values $\theta_1, \theta_2, \dots, \theta_m$ under which data sets are simulated from P_θ . After obtaining $\mathcal{Z}_{n,m}$, for any new parameter value θ from the prior distribution under which we want to simulate a new data set, we substitute θ with $\tilde{\theta}$ such that $(\tilde{\mathbf{x}}, \tilde{\theta}) \in \mathcal{Z}_{n,m}$. The value $\tilde{\theta}$ has the minimum Euclidean distance to the value θ among all parameter values in $\mathcal{Z}_{n,m}$. More precisely, $\tilde{\theta} = \arg \min_{\theta_j \in \mathcal{Z}_{n,m}} \|\theta_j - \theta\|$. In essence, this approximation is

equivalent to replacing the sampling distribution of the data set P_θ with the sampling distribution $P_{\tilde{\theta}}$; we call this an approximation on the parameter space. However, this parameter space approximation is not sufficient to simulate data sets efficiently, since the model $P_{\tilde{\theta}}$ is still limited-generative after this substitution.

As a second approximation, we substitute the model $P_{\tilde{\theta}}$ with the empirical

distribution of the data set $\tilde{\mathbf{x}}$ that has already been simulated from $P_{\tilde{\theta}}$ as $(\tilde{\mathbf{x}}, \tilde{\theta}) \in \mathcal{Z}_{n,m}$. Here, we assume a positive probability mass only on the data values observed in the set $\tilde{\mathbf{x}}$. We call this an approximation on the model space because the model $P_{\tilde{\theta}}$ is substituted with the empirical distribution of a data set simulated from $P_{\tilde{\theta}}$.

To simulate a new data set \mathbf{x} in AABC, we utilize a vector of positive auxiliary parameters $\phi = (\phi_1, \phi_2, \dots, \phi_n)$, that satisfy $\sum_{i=1}^n \phi_i = 1$. We let ϕ_i be the probability that a random data value $x_j \in \mathbf{x}$ is equal to a given value \tilde{x}_i found in the data set $\tilde{\mathbf{x}} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$. The premise is that the sample $\tilde{\mathbf{x}}$ simulated under $\tilde{\theta}$ provides information about the model $P_{\tilde{\theta}}$, and by an approximation of θ to $\tilde{\theta}$ on the parameter space, about P_{θ} .

If we denote the approximate sampling distribution of a data set $\mathbf{x} = (x_1, x_2, \dots, x_n)$ by Q_{θ} , its joint probability mass function is

$$\int_{\Phi} q(\mathbf{x}|\phi, \tilde{\mathbf{x}}) \pi(\phi) d\phi \mathbf{I}_{\{\theta, \tilde{\theta}\}}, \quad (3)$$

where $q(\mathbf{x}|\phi, \tilde{\mathbf{x}}) = \binom{n}{n_1 \ n_2 \ \dots \ n_k} \prod_{j=1}^n \prod_{i=1}^n \phi_i^{\mathbf{I}_{\{x_j = \tilde{x}_i\}}}$, and $\mathbf{I}_{\{\theta, \tilde{\theta}\}}$ is 1 if $\tilde{\theta} \in \mathcal{Z}_{n,m}$ is the closest value to θ in the Euclidean sense and is 0 otherwise. Here, n_i is the number of times \tilde{x}_i observed in the new sample \mathbf{x} , k is the number of distinct data values observed in the data set \mathbf{x} , and $\mathbf{I}_{\{x_j = \tilde{x}_i\}}$ is 1 if $x_j = \tilde{x}_i$ and is 0 otherwise. The distribution $q(\mathbf{x}|\phi, \tilde{\mathbf{x}})$ is that of an IID sample $\mathbf{x} = (x_1, x_2, \dots, x_n)$, where x_j is drawn from the values $(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$ with probabilities $(\phi_1, \phi_2, \dots, \phi_n)$.

The probability vector ϕ is a parameter of the model conditional on $\tilde{\mathbf{x}}$, and thus, we need to posit a prior distribution on ϕ . As a natural prior on probabilities, we let the prior distribution $\pi(\phi)$ on ϕ be the symmetric Dirichlet distribution on the $(n-1)$ -dimensional simplex Φ , with hyperparameters $(1, 1, \dots, 1)$ and a uniform probability density function proportional to 1. This choice assigns equal weight to all distributions placing positive probability mass on the data points $\tilde{x}_i \in \tilde{\mathbf{x}}$. Further, it assigns zero posterior probability to data values unobserved in the sample $\tilde{\mathbf{x}}$, thereby avoiding difficulties created by such values in the likelihood (Rubin, 1981; Owen, 1990).

To distinguish the parameter and data set realizations in $\mathcal{Z}_{n,m} = \{(\mathbf{x}_i, \theta_i)\}_{i=1}^m$ from the parameter and data sets simulated using AABC, we use starred versions of each quantity to denote specific values simulated in AABC. For example, as the sampling distribution P_{θ_i} delivers a data set \mathbf{x}_i under a given parameter value θ_i in the ABC procedure of Algorithm 2, the sampling distri-

Notation		Used in	Description
Random	Realized		
θ	θ_i	ABC/AABC	Parameter value.
	θ_i^*	AABC	Parameter value.
$\tilde{\theta}$	$\tilde{\theta}_i$	AABC	Parameter value in the set $\mathcal{Z}_{n,m}$ closest to θ or θ_i^* .
ϕ	ϕ_i	AABC	Auxiliary parameter value for θ or θ_i^* .
	ϕ_{ij}	AABC	j th element of ϕ_i .
\mathbf{x}	\mathbf{x}_i	ABC	Data set of n IID observations generated from P_θ or P_{θ_i} .
	\mathbf{x}_i^*	AABC	Data set of n IID observations generated from $Q_{\theta_i^*}$.
$\tilde{\mathbf{x}}$	$\tilde{\mathbf{x}}_i$	AABC	Data set in $\mathcal{Z}_{n,m}$ simulated from $P_{\tilde{\theta}}$ or $P_{\tilde{\theta}_i}$.
x_j	x_{ij}	ABC	j th data point in the data set \mathbf{x} or \mathbf{x}_i .
	x_{ij}^*	AABC	j th data point in the data set \mathbf{x}_i^* .
\tilde{x}_j	\tilde{x}_{ij}	AABC	j th data point in the data set $\tilde{\mathbf{x}}$ or $\tilde{\mathbf{x}}_i$.

Figure 4: Notation used in the text and algorithms.

bution $Q_{\theta_i^*}$ delivers a data set \mathbf{x}_i^* under a given parameter value θ_i^* simulated from its prior distribution (see Figure 4 for notation).

The sampling distribution Q_θ utilizes the information available in the set of realizations $\mathcal{Z}_{n,m}$ through the parameter ϕ , since the prior distribution of ϕ conditions on $(\tilde{\mathbf{x}}, \tilde{\theta}) \in \mathcal{Z}_{n,m}$ and thus on the set $\mathcal{Z}_{n,m}$. In this sense, the available realizations $\mathcal{Z}_{n,m}$ are used as fixed background information about P_θ , and inferences using the substitute model Q_θ are conditional on the simulated sets $\mathcal{Z}_{n,m}$.

3.2 The posterior distribution of θ sampled by AABC

In sampling the approximate posterior distribution of θ by AABC methods, we use the two ABC approximations described in Section 2. First, we substitute each data instance \mathbf{x} with summary statistics \mathbf{s} . Second, we use an acceptance condition with tolerance ϵ , employing the Euclidean distance to measure the proximity of the summary statistics calculated from the observed and simulated data, as in equation 1. If we let θ_j^* be a new parameter value simulated from its prior distribution after obtaining the set $\mathcal{Z}_{n,m}$, in AABC we accept the parameter values θ_j^* producing summary statistics \mathbf{s}_j^* that satisfy the condition $\|\mathbf{s}_j^* - \mathbf{s}_o\| < \epsilon$ as being draws from the posterior

distribution. This acceptance condition corresponds to a uniform kernel, which we use throughout this article, although like ABC, AABC can employ other kernels to obtain smooth weighting of \mathbf{s}_j^* values by their distance from \mathbf{s}_o . Substituting P_θ with Q_θ involves replacing $p(\mathbf{x}|\theta)$ in expression 2 with expression 3 and adjusting the normalizing constant accordingly. The approximate posterior distribution sampled by an AABC method is

$$\pi_\epsilon(\theta|\mathbf{x}_o, Q_\theta) = \frac{1}{C_{Q_\theta}} \int_{\mathcal{X}} \mathbf{I}_{\{\|\mathbf{s}-\mathbf{s}_o\|<\epsilon\}} \left[\int_{\Phi} q(\mathbf{x}|\phi, \tilde{\mathbf{x}}) \pi(\phi) d\phi \mathbf{I}_{\{\theta, \tilde{\theta}\}} \right] \pi(\theta) d\mathbf{x}, \quad (4)$$

where $C_{Q_\theta} = \int_{\Theta} \int_{\mathcal{X}} \mathbf{I}_{\{\|\mathbf{s}-\mathbf{s}_o\|<\epsilon\}} \left[\int_{\Phi} q(\mathbf{x}|\phi, \tilde{\mathbf{x}}) \pi(\phi) d\phi \mathbf{I}_{\{\theta, \tilde{\theta}\}} \right] \pi(\theta) d\mathbf{x} d\theta$ is the normalizing constant.

The AABC approach is sensible in that as the limited generative model increasingly permits a larger number of simulated data sets, for large sample sizes the posterior distribution obtained by an AABC method approaches the same distribution as the posterior distribution obtained by an ABC method. We codify this claim with a theorem.

Theorem. Let $\pi(\theta)$ be a bounded prior on θ . Let $\pi_\epsilon(\theta|\mathbf{x}_o, P_\theta)$ and $\pi_\epsilon(\theta|\mathbf{x}_o, Q_\theta)$ be the posterior distributions sampled by a standard ABC method and an AABC method, respectively. Then

$$\lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \pi_\epsilon(\theta|\mathbf{x}_o, Q_\theta) = \lim_{n \rightarrow \infty} \pi_\epsilon(\theta|\mathbf{x}_o, P_\theta). \quad (5)$$

A proof of the theorem is given in Appendix 1. The convergence of the posterior distribution sampled by AABC is a consequence of the fact that, for each given value of θ , the sampling distribution $\int_{\Phi} q(\mathbf{x}|\phi, \tilde{\mathbf{x}}) \pi(\phi) d\phi \mathbf{I}_{\{\theta, \tilde{\theta}\}}$ converges to the true sampling distribution $p(\mathbf{x}|\theta)$ as the sample size n and the number of simulated samples m from P_θ increase. The intuition for the double limit in equation 5 is as follows. The standard notion of a distribution converging to a point in the parameter space as the sample size n increases does not directly apply to the posterior distribution $\pi_\epsilon(\theta|\mathbf{x}_o, Q_\theta)$, since this posterior depends not only on the sample size n , but also on the number m of simulated data sets from P_θ . Hence, for convergence of the posterior distribution based on the likelihood of Q_θ , the requirement is that both $n \rightarrow \infty$ and $m \rightarrow \infty$. As $n \rightarrow \infty$, the empirical distribution converges to $P_{\tilde{\theta}}$, the correct sampling distribution with the incorrect parameter value $\tilde{\theta}$. As $m \rightarrow \infty$, the distance between the parameter value θ under which we want

to simulate a new data set and the parameter value $\tilde{\theta} \in \mathcal{Z}_{n,m}$ closest to θ approaches zero. Therefore, taking both limits simultaneously results in convergence to the correct sampling distribution P_θ .

3.3 AABC algorithms

The structure of AABC algorithms sampling the posterior distribution in expression 4 can be conveniently summarized in three parts, as shown in AABC by a rejection algorithm (Figure 5). In Algorithm 2, Part I involves obtaining a limited number of realizations from the joint distribution of the parameter and the data from the limited-generative model P_θ . Part I simply involves the application of steps 1 and 2 from Algorithm 1, but only for m iterations. Part II involves simulating a new parameter value θ_i^* from its prior distribution (step 4) and then simulating a data set \mathbf{x}_i^* from the model $Q_{\theta_i^*}$ (steps 5, 6, 7), conditional on $\mathcal{Z}_{n,m}$ obtained in Part I. Part III involves comparing the summary statistics \mathbf{s}_i^* calculated from the simulated data set \mathbf{x}_i^* with the summary statistics \mathbf{s}_o calculated from the observed data set \mathbf{x}_o , to accept or reject the parameter value θ_i^* . The calculation and comparison of summary statistics follows the same procedure as in steps 3 and 4 of Algorithm 1. Hence, Part II of AABC by rejection has the novel steps 5, 6, and 7, whereas Parts I and III use the machinery of ABC by rejection from Algorithm 1.

We can show that Algorithm 2 samples the correct posterior distribution $\pi_\epsilon(\theta|\mathbf{x}_o, Q_\theta)$. The probability of sampling a parameter value θ in Algorithm 2 is proportional to

$$\begin{aligned} & \sum_{\mathbf{s}} \sum_{\phi} \pi(\theta) \mathbf{I}_{\{\theta, \tilde{\theta}\}} \pi(\phi) q(\mathbf{x}|\phi, \tilde{\mathbf{x}}) \mathbf{I}_{\{\|\mathbf{s} - \mathbf{s}_o\| < \epsilon\}} \\ &= \sum_{\mathbf{s}} \sum_{\phi} \pi(\theta, \phi) \mathbf{I}_{\{\theta, \tilde{\theta}\}} q(\mathbf{x}|\phi, \tilde{\mathbf{x}}) \mathbf{I}_{\{\|\mathbf{s} - \mathbf{s}_o\| < \epsilon\}} \\ &\propto \sum_{\mathbf{s}} \sum_{\phi} \pi(\theta, \phi|Q_\theta) \mathbf{I}_{\{\|\mathbf{s} - \mathbf{s}_o\| < \epsilon\}} \\ &\propto \pi_\epsilon(\theta|\mathbf{x}_o, Q_\theta), \end{aligned}$$

where the third line follows from the fact that the expression on the second line is the product of the likelihood under the model Q_θ and the prior, and therefore it is proportional to the posterior distribution of parameters based on the model Q_θ .

Algorithm 2: AABC by rejection algorithm

Part I		1. For $i = 1, 2, \dots, m$, simulate $\theta_i \sim \pi(\theta)$.
		2. For $i = 1, 2, \dots, m$, simulate $\mathbf{x}_i \sim P_{\theta_i}$.
		3. Let $\mathcal{Z}_{n,m} = \{(\mathbf{x}_1, \theta_1), \dots, (\mathbf{x}_m, \theta_m)\}$.
Part II		4. Simulate $\theta_i^* \sim \pi(\theta)$.
		5. Find $\tilde{\mathbf{x}}_i = (\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{in})$, where $(\tilde{\mathbf{x}}_i, \tilde{\theta}_i) \in \mathcal{Z}_{n,m}$ and $\tilde{\theta}_i$ is the closest value to θ_i^* .
		6. Simulate $\phi_i \sim \pi(\phi)$.
Part III		7. Simulate a new data set $\mathbf{x}_i^* = (x_{i1}^*, x_{i2}^*, \dots, x_{in}^*)$, where x_{ij}^* are IID from values $(\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{in})$ with the probability of \tilde{x}_{ij} proportional to ϕ_{ij} .
		8. Calculate the summary statistics \mathbf{s}_i^* from \mathbf{x}_i^* .
		9. If $\ \mathbf{s}_o - \mathbf{s}_i^*\ < \epsilon$, output θ_i^* .

Figure 5: The AABC algorithm by rejection sampling. One iteration of the algorithm is shown on the right, along with a schematic illustration of sampling from the posterior distribution of θ based on M proposed parameter values in the rejection algorithm (left).

4 Applications

In this section, we investigate the inferential performance of AABC approach with two examples. The following simulation setup is used in both examples.

4.1 Simulation study design

We simulated a reference set with $M = 10^5$ realizations $\{(\mathbf{x}_1, \theta_1), (\mathbf{x}_2, \theta_2), \dots, (\mathbf{x}_{10^5}, \theta_{10^5})\}$, by first generating $\theta_i \sim \pi(\theta)$ and then simulating a data set $\mathbf{x}_i \sim P_{\theta_i}$. We then sampled 1000 pairs (\mathbf{x}_i, θ_i) from the reference set, uniformly at random without replacement. Thus, we selected 1000 “true” parameter values θ_i , along with corresponding test data sets \mathbf{x}_i generated under each value θ_i from the model P_{θ_i} . Further, we built the sets $\mathcal{Z}_{n,m}$, with $m = 10^2, 5 \times 10^2, 10^3, 5 \times 10^3, 10^4, 5 \times 10^4, 10^5$ by sampling the reference set uniformly at random without replacement for $m < 10^5$, and taking all the real-

izations in the reference set for $m = M = 10^5$. The sample size n of the data is described in each relevant example.

On each test data set, we performed AABC by rejection sampling (Algorithm 2) using each set $\mathcal{Z}_{n,m}$. In example 1, where our goal is to compare the performance of the AABC and ABC approaches, we performed ABC analyses by rejection sampling (Algorithm 1) using the same sets $\mathcal{Z}_{n,m}$. For all analyses, we obtained a sample from the joint posterior distribution of the parameter vector θ by accepting the parameter vector values that generated data whose summary statistics were in the top 1 percentile with respect to the statistics calculated from the test data set, in the sense of equation 1. Compared to the approach of fixing the ϵ cutoff, accepting parameter vectors that generate data whose summary statistics are in a top percentile has the advantage that a desired number of samples from the posterior is always obtained given a total fixed number of proposed parameter values. This approach is often preferred by ABC practitioners and is convenient in our case for comparing ABC and AABC.

We assessed the accuracy of the posterior samples for each component of the parameter vector θ separately, using the root sum of squared error for standardized parameter values accepted in the posterior sample. For a generic scalar parameter α , the root sum of squared errors is given by $\text{RSSE} = (1/r)\sqrt{\sum_{j=1}^r (\alpha_j - \alpha_T)^2 / \text{Var}(\alpha)}$, where $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_r)$ are r accepted values in the posterior sample, α_T is the true parameter value, and $\text{Var}(\alpha)$ is the variance of the set of r values. We report the mean RSSE over 1000 test data sets as $\text{RMSE} = (1/1000) \sum_{i=1}^{1000} \text{RSSE}_i$ (see Nunes and Balding (2010)).

4.2 Example 1: The strength of balancing selection in a multi-locus K -allele model

In this section, we consider inference from the stationary distribution of allele frequencies in the diffusion approximation to a Wright-Fisher model with symmetric balancing selection and mutation (Wright, 1949). If we let $a_i > 0$, with $i = 1, 2, \dots, K$, and $\sum_{i=1}^K a_i = 1$, and denote the frequency of allelic type i in the population at a genetic locus, the joint probability density function of allele frequencies $x = (a_1, a_2, \dots, a_K)$ is $f(x|\sigma, \mu) = c(\sigma, \mu)^{-1} \exp(-\sigma \sum_{i=1}^K a_i^2) \prod_{i=1}^K a_i^{\mu/K-1}$. Parameters σ and μ determine the population-scaled strength of balancing selection and the mutation rate, re-

spectively. A data set of observed allele frequencies is a random sample of n draws from the population frequencies $f(x|\sigma, \mu)$.

ABC methods are well-suited for inference from this model for three reasons. First, the statistics $\sum_{j=1}^K a_j^2$ and $-\sum_{j=1}^K \log a_j$ are jointly sufficient for parameters σ and μ , and no information loss occurs in dimension reduction to the summary statistics. Second, the parameter-dependent normalizing constant $c(\sigma, \mu)$ is hard to calculate, and performing likelihood-based inference on σ and μ is therefore difficult. Third, a method specifically designed to simulate data sets from $f(x|\sigma, \mu)$ is readily available (Joyce *et al.*, 2012), and performing ABC is therefore straightforward. For simplicity, we assume 100 loci with the same true parameter values, each with $K = 4$, and that the allele frequencies at each locus are independent of the allele frequencies at other loci. Thus, the joint probability density function of allele frequencies for 100 loci is equal to the product of probability density functions across loci. We choose uniform prior distributions, on $(0.1, 10)$ for the mutation rate (μ), and on $(0, 50)$ for the selection parameter (σ).

Results. Posterior samples model parameters (σ, μ) obtained by ABC and AABC using a typical data set are given in Figure 6. In analyses with $m = 10^2, 5 \times 10^2, 10^3$ or 5×10^3 simulated data sets, few samples are accepted with ABC, and thus, little mass is observed in ABC histograms (black). For small m , ABC does not produce an adequate sample size from the posterior distribution of parameters. AABC, however, produces a posterior sample of size 10^3 for any m , because 10^5 data sets are simulated from the non-mechanistic model (Algorithm 2, steps 5, 6, 7) and the top 1 percentile are accepted as belonging to the approximate posterior distribution. The histograms obtained by AABC recover the true value reasonably well (Figure 6). The RMSE values in AABC procedures are approximately constant with increasing m . For $m = 10^2, 5 \times 10^2, 10^3, 5 \times 10^3, 10^4, 5 \times 10^4$, and 10^5 simulated data sets, the RMSE values for parameter μ are 5.988, 5.932, 6.012, 6.086, 6.125, 6.078, and 6.088 respectively, close to the RMSE of 5.290 obtained by a standard ABC approach using $M = 10^5$ simulated data sets from the mechanistic model. The RMSE values in the last column of Figure 6 show that an AABC approach produces posterior samples that have on average greater variance than posterior samples obtained from ABC with the same large number of realizations. Here, greater variance in posterior samples obtained by AABC is a result of simulating data sets in AABC by resampling the observed data values that are found only in the m realizations in $\mathcal{Z}_{n,m}$. Consider two parameter values θ_1^* and θ_2^* for which data sets \mathbf{x}_1^* and \mathbf{x}_2^* are

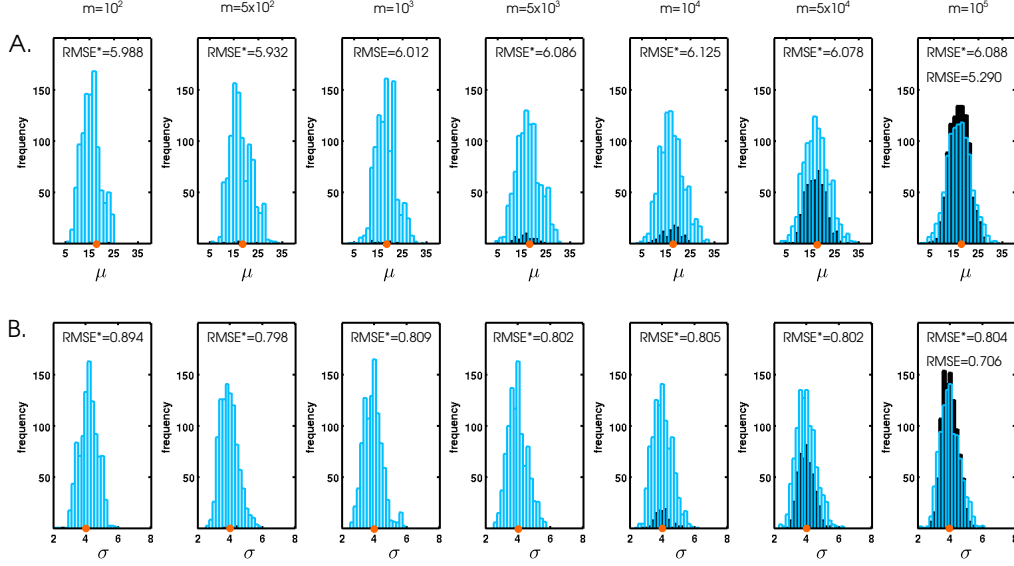


Figure 6: Inference on the strength of balancing selection. The figure shows the marginal posterior distributions of parameters μ (A), and σ (B) of example 1 obtained with ABC by rejection (black) and with AABC by rejection (blue). The number m of data sets simulated from the mechanistic model for each analysis performed by AABC and ABC appears at the top of each column. The red dot on the x-axis is the true value of the parameter, equal in all plots. RMSE* values in each plot are from AABC analyses, averaged over 1000 test data sets. RMSE values in the last column are from corresponding ABC analyses.

simulated in the AABC approach by steps 5, 6, 7 of Algorithm 2 such that the parameter value $\tilde{\theta} \in \mathcal{Z}_{n,m}$ closest to both θ_1^* and θ_2^* is the same value. The data sets \mathbf{x}_1^* and \mathbf{x}_2^* can include only the data values observed in $\tilde{\mathbf{x}}$ of the pair $(\tilde{\mathbf{x}}, \tilde{\theta}) \in \mathcal{Z}_{n,m}$. On average, \mathbf{x}_1^* and \mathbf{x}_2^* share more observations in common than two data sets simulated from the respective mechanistic models $P_{\theta_1^*}$ and $P_{\theta_2^*}$. Therefore, each data set simulated in the AABC approach using Q_θ is expected to be less able to distinguish between different parameter values than the independent data sets simulated in the ABC approach using P_θ . This situation results in relatively flat likelihoods and hence posterior samples with larger variance.

4.3 Example 2: Admixture rates in hybrid populations

Models in which hybrid populations are founded by, and receive genetic contributions from, multiple source populations are of interest in describing the demographic history of admixture. Stochastic models including admixture often result in likelihoods that are difficult to calculate, and statistical methods capable of performing inference on admixture rates have received much attention for their implications on topics ranging from human evolution to conservation ecology (Falush *et al.*, 2003; Tang *et al.*, 2005; Buerkle and Lexer, 2008). Here, we consider inference on admixture rates from a mechanistic model of Verdu and Rosenberg (2011). We use reported estimates of individual admixture as data.

We consider a model of admixture for a diploid hybrid population of constant size N , founded at some known t generations in the past with contributions from source populations A and B. We follow the distribution of admixture fractions of individuals in the hybrid population at a given genetic locus. Each generation, the admixture fraction for each individual in the hybrid population is obtained as the mean of the admixture fractions of its parents. The parents are chosen independently of each other, from source population A, source population B, or the hybrid population of the previous generation with probabilities p_A, p_B , and p_H , respectively ($p_A + p_B + p_H = 1$). In the special case of the founding generation, $p_H = 0$, and we assume $p_A = p_B = 0.5$. Individuals from source populations A and B are assigned admixture fractions of 1 and 0 respectively. For example, if both parents of an individual in the hybrid population of the founding generation are from source population A, that individual has admixture fraction $(1 + 1)/2 = 1$. If both parents are from population 2, the admixture fraction is $(0 + 0)/2 = 0$, and if one parent is from population 1 and the other is from population B, then the admixture fraction is $(1 + 0)/2 = 0.5$. The distribution of the admixture fraction in the hybrid population is propagated in this manner for t generations until the present, in which a sample of n individuals is obtained from the resulting distribution (Figure 7). Our goal is to estimate the admixture rates (p_A, p_B, p_H) , given the individual admixture fractions estimated from observed genetic data.

We apply the AABC approach using individual admixture fractions from $n = 604$ individuals from Central African Pygmy populations reported by Verdu *et al.* (2009), with an assumed constant population size of $N = 10^4$. This assumption differs slightly from the original model in Verdu and Rosen-

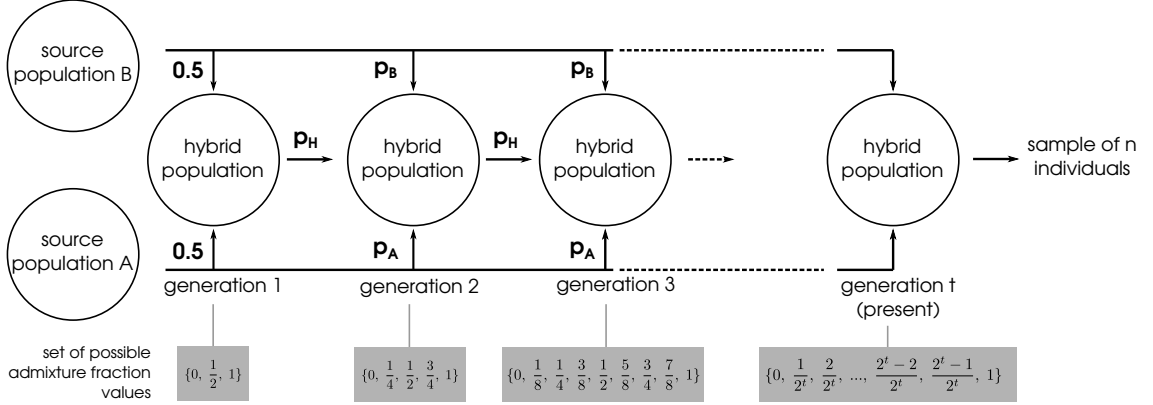


Figure 7: The admixture model of example 2.

berg (2011) in that a finite population size is assumed, so that only 10^4 admixture fraction values are allowed in the population at any given generation. We assume that an admixture event with contributions from two ancestral source populations started at the mean estimate of $t = 771$ generations ago (Verdu *et al.*, 2009) with a generation time of 25 years, and that it continued until the present. Source population A refers to an ancestral Pygmy population, and source population B refers to an ancestral non-Pygmy population. The feature of this model relevant to our method is the computational intractability of simulating data sets. For each set of parameter values (p_A, p_B, p_H) simulated from the priors, the distribution of admixture fractions is discrete on a support of a number of admixture fraction values that doubles each generation, and this distribution evolves for 771 generations. A random sample of admixture fraction values comparable to the values calculated from the observed data set is obtained from the distribution of the present generation. Simulating a large number of data sets under this model with such a large number of generations is computationally infeasible, and standard ABC is impractical. We thus perform AABC by rejection (Algorithm 2) using $m = 10^4$ realizations from this model. We assume a Dirichlet prior with hyperparameters $(1, 1, 1)$ on parameters (p_A, p_B, p_H) .

We also assessed the contribution of the approximations on the parameter and model spaces in the AABC approach to the RMSE separately, with a simulation study using a small number of generations ($t = 30$), where simulating data sets from the mechanistic model is feasible. First, we per-

formed AABC with rejection as in Algorithm 2 with 1000 “true” data sets using $m = 10^2, 5 \times 10^2, 10^3, 5 \times 10^3, 10^4, 5 \times 10^4$, and 10^5 realizations from the model, and we calculated the RMSE for p_A, p_B , and p_H over 1000 “true” data sets as described in Section 4.1. This AABC analysis includes error due to approximations on the parameter space and on the model space. Second, we performed an AABC analysis with the same set of m realizations, by including the error only due to the approximation on the parameter space. We achieved this by running Algorithm 2 up through step 5, and then simulating data sets from the mechanistic model by substituting steps 6 and 7 of Algorithm 2 with step 2 of Algorithm 1, the standard ABC approach by rejection. By this substitution, all data sets are simulated from the mechanistic model, but each data set is obtained using a parameter vector $(\tilde{p}_A, \tilde{p}_B, \tilde{p}_H)$ found in step 5 of Algorithm 2. In this procedure, the error due to the approximation on the model space is eliminated, because data sets are simulated from the correct mechanistic model and not by resampling from the available realizations in $\mathcal{Z}_{n,m}$. However, this procedure includes error due to the approximation on the parameter space, because each data set is simulated not under the correct proposed parameter value, but under the parameter value $(\tilde{p}_A, \tilde{p}_B, \tilde{p}_H)$, the closest value to the correct proposed value that can be found in $\mathcal{Z}_{n,m}$. We compared the RMSE of the AABC procedure involving the approximation on both the parameter and model spaces and the RMSE of the AABC procedure involving only the approximation on the parameter space to the RMSE obtained from a standard ABC approach. For these two AABC procedures, we also compared the percent excess in RMSE, defined as the ratio of the absolute difference in RMSE of the AABC and standard ABC approaches to the RMSE of the standard ABC approach, expressed as a percent.

Results. The individual admixture fractions calculated from the Pygmy data carry substantial information about the admixture parameters p_A, p_B , and p_H , since the joint posterior distribution is concentrated in a relatively small region of the 3-dimensional unit simplex on which (p_A, p_B, p_H) sits (Figure 8A). The marginal posterior distributions (Figure 8B, 8C, and 8D) have means $p_A = 0.151$, $p_B = 0.132$, and $p_H = 0.717$. These values are interpreted as contribution of genetic material of 15.1% from the ancestral Pygmy population (source population A), 13.2% from the ancestral Non-Pygmy population (source population B), and 71.7% from the hybrid population to itself at each generation, over 771 generations of constant admixture.

For the simulation study with $t = 30$ generations and 1000 “true data”

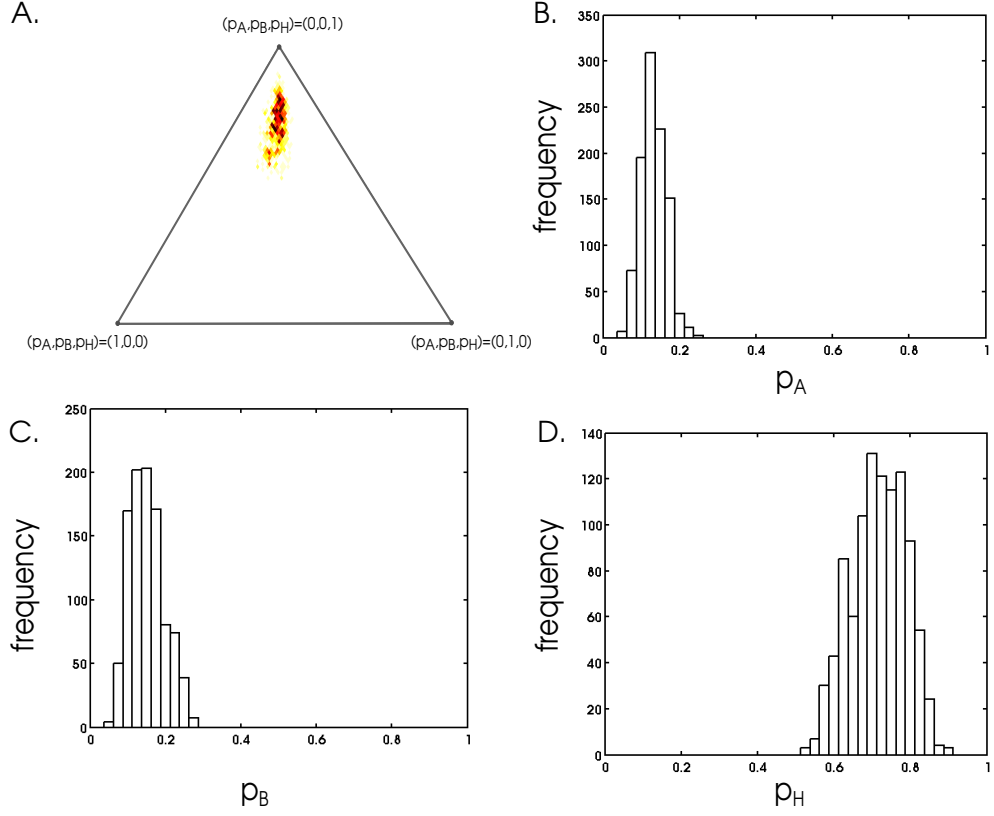


Figure 8: AABC analysis on the Pygmy data of example 2 with $m = 10^4$ realizations under the mechanistic model. (A) The joint distribution on the unit simplex, with probability mass increasing from white to dark red. (B,C,D) Marginal distributions of p_A , p_B , and p_H .

sets, the RMSE values from AABC analyses decrease with increasing m (Figure 9A, 9B, 9C). Further, as m increases, the error due to the approximation on the parameter space decreases (Figure 9D last column), due to the fact that for large m , the difference decreases between the closest parameter value chosen at step 5 of Algorithm 2 and the correct parameter value under which we want to simulate a data set. In fact, the RMSE from the AABC analysis with $m = 10^5$ realizations and approximation only on the parameter space and the RMSE from the standard ABC approach are virtually indistinguishable (Figure 9A, 9B, 9C, red star). For $m = 10^3$, the AABC analysis with approximations on the parameter and model spaces has a percent excess RMSE of 13.81%, whereas AABC analysis including only the approximation on the parameter space has excess RMSE of 6.61%. That is, at $m = 10^3$, approximately half of the excess RMSE in the AABC approach with respect to the standard ABC analysis comes from the error due to the approximation on the parameter space and half arises due to the approximation on the model space.

5 Discussion

Performing likelihood-based inference from statistical models incorporating a multitude of stochastic processes is often challenging due to computationally intractable likelihoods. In principle, when stochastic processes are complex but a family of parametric statistical models is well-defined, data can be simulated from the model to assess the parameter likelihoods. In the last decade, ABC methods have become a standard tool to perform approximate Bayesian inference in subject areas such as ecology and evolution, by exploiting the idea of simulating many data sets from a model, when such simulations are computationally feasible. To deliver an adequate sample from the posterior distribution of the parameters, however, ABC requires a large number of simulated data sets, and it might not perform well when only a limited number of data sets can be simulated.

In this article, we introduced an approach that extends simulation-based Bayesian inference methods to model spaces in which only a limited number of data sets can be simulated from the model, at the expense of requiring approximations on the parameter and the model spaces. Our AABC approaches rely on two statistical approximations. In our approximation on the parameter space, for each parameter simulated from the prior distribu-

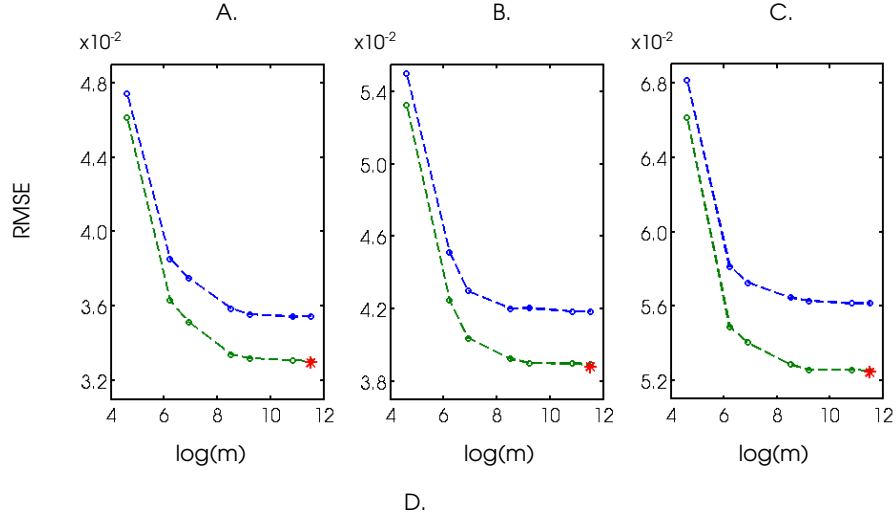


Figure 9: RMSE in the admixture model. The decrease in RMSE is shown for parameters p_A (A), p_B (B), and p_H (C) with increasing m , the number of simulated samples from the mechanistic model, for AABC analysis performed with an approximation only on the parameter space (green), and with an approximation on both the parameter space and the model space (blue). The red star in each plot is the RMSE obtained by a standard ABC analysis performed with $M = 10^5$ simulated values. (D) The percent excess in RMSE of the two AABC approaches relative to a standard ABC approach for parameters p_A , p_B , and p_H .

tion, we take the closest parameter value available in the set of realizations $\mathcal{Z}_{n,m}$ obtained from the mechanistic model. This approach has a uniform kernel smoothing interpretation in the sense that each parameter value in the set $\mathcal{Z}_{n,m}$ dissects the support of the prior distribution into non-overlapping components such that each interval is mapped to the same parameter value in $\mathcal{Z}_{n,m}$. Each component then represents the support of a uniform kernel. Kernel approximations have an operational role in implementing ABC methods, and a natural future direction for AABC is to improve the accuracy of posterior samples using smooth weighting kernels for the approximation on the parameter space.

The approximation on the model space is achieved by assigning Dirichlet probabilities to data points of realizations obtained from the mechanistic model. This is a variation on the resampling method originally introduced in Rubin’s Bayesian bootstrap (Rubin, 1981), and therefore, it is an application of Bayesian nonparametric methods. From this perspective, AABC methods connect standard model-based Bayesian inference on model-specific parameters and Bayesian nonparametric methods within the ABC framework.

Our approach of using a non-mechanistic model and Bayesian resampling methods to help perform inference on model-specific parameters of a mechanistic model is a fundamental difference between AABC and existing ABC methods. ABC performs inference on model-specific parameters of a mechanistic model using a likelihood based purely on the mechanistic model. AABC instead performs inference on the same model-specific parameters of the mechanistic model as ABC, using a likelihood based on a non-mechanistic model that incorporates a limited number of data sets simulated from the mechanistic model. Consequently, the model likelihoods used in ABC and AABC are not exactly the same, and the posterior distributions targeted by the two classes of methods are not exactly equivalent for finite sample sizes. The advantage of AABC methods in contrast to pure non-mechanistic modeling approaches (e.g., nonparametric methods) is that AABC can perform inference on the quantities of interest—the model-specific parameters of the mechanistic model.

Unlike other ABC methods, the AABC approach delivers a posterior sample of desired size from the joint distribution of parameters for any $m > 1$. This is both a strength and a limitation of AABC. The strength is that in practice, a researcher can fix m and thus the computation time *a priori*, to simulate data from the mechanistic model to obtain a reasonable inference by AABC; other ABC methods may fail to produce an adequate posterior

sample in equivalent computation time. In our example, for moderate values of m (e.g., 10^3 to 10^4) for which standard ABC approaches were unsatisfactory, AABC adequately sampled an approximate posterior distribution. The limitation is that when m is too small, the posterior sample obtained by AABC can be a distorted representation of the true posterior distribution. Although in the limit, AABC and ABC are expected to produce similar results, the posterior distribution sampled by an AABC approach is not the correct posterior distribution, because many parameter values simulated from the prior are tested for acceptance based on repeated use of the data values in m realizations, instead of based on data sets simulated independently of each other. A future direction is to investigate the relationship between m and the dimensionality of the parameter space to optimize m in producing a given level of accuracy for approximating the true posterior distributions.

Acknowledgments

The authors thank Paul Verdu for helpful discussions on the genetics of Central African Pygmy populations. Support for this research is partially provided by NIH grant R01 GM 081441, NSF grant DBI-1146722, and the Burroughs Wellcome Fund.

References

- Aeschbacher, S., Futschik, A., and Beaumont, M.A. (2012), “A novel approach for choosing summary statistics in approximate Bayesian computation,” *Genetics*, 192, 1027–1047.
- Beaumont, M.A. (2003), “Estimation of population growth or decline in genetically monitored populations,” *Genetics*, 164, 1139–1160.
- Beaumont, M.A. (2010), “Approximate Bayesian computation in evolution and ecology,” *Annual Reviews of Ecology, Evolution and Systematics*, 41, 379–406.
- Beaumont, M.A., Zhang, W., and Balding, D.J. (2002), “Approximate Bayesian computation in population genetics,” *Genetics*, 162, 2025–2035.

- Beaumont, M.A., Cornuet, J.-M., Marin, J.-M., and Robert, C.P. (2009), “Adaptive approximate Bayesian computation,” *Biometrika*, 96, 983–990.
- Becquet, C., and Przeworski, M. (2007), “A new approach to estimate parameters of speciation models with application to apes,” *Genome Research*, 17, 1505–1519.
- Blum, M.G.B., and François, O. (2010), “Non-linear regression models for approximate Bayesian computation,” *Statistics and Computing*, 20, 63–73.
- Blum, M.G.B., and Jakobsson, M. (2010), “Deep divergences of human gene trees and models of human origins,” *Molecular Biology and Evolution*, 28, 889–898.
- Bortot, P., Coles, S., and Sisson, S.A. (2007), “Inference for stereological extremes,” *Journal of the American Statistical Association*, 102, 84–92.
- Buerkle, C.A. and Lexer, C. (2008), “Admixture as the basis for genetic mapping,” *Trends in Ecology and Evolution*, 23, 686–694.
- Cornuet, J.-M., Santos, F., Beaumont, M.A., Robert, C.P., Marin, J.-M., Balding, D.J., Guillemaud, T., and Estoup, A. (2008), “Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation,” *Bioinformatics*, 24, 2713–2719.
- Csilléry, K., Blum, M.G.B., Gaggiotti, O.E., and François, O. (2010), “Approximate Bayesian computation (ABC) in practice,” *Trends in Ecology and Evolution*, 25, 410–418.
- Diggle, P.J., and Gratton, R.J. (1984), “Monte Carlo methods of inference for implicit statistical models,” *Journal of the Royal Statistical Society, Series B* 46, 193–227.
- Estoup, A., Beaumont, M.A., Sennedot, F., Moritz, C., and Cornuet, J.-M. (2004), “Genetic analysis of complex demographic scenarios: spatially expanding populations of the cane toad, *Bufo marinus*,” *Evolution* 58, 2021–2036.
- Fagundes, N.J.R., Ray, N., Beaumont, M.A., Neuenschwander, S., Salzano, F.M., Bonatto, S.L., and Excoffier, L. (2007), “Statistical evaluation of alternative models of human evolution,” *Proceedings of the National Academy of Sciences*, 104, 17614–17619.

- Falush, D., Stephens, M., and Pritchard, J.K. (2003), “Inference of population structure using multilocus genetic data: linked loci and correlated allele frequencies,” *Genetics*, 164, 1567–1587.
- Fearnhead, P., and Prangle, D. (2012), “Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation,” *Journal of the Royal Statistical Society, Series B*, 74, 1–28.
- François, O., Blum, M.G.B., Jakobsson, M., and Rosenberg, N.A. (2008), “Demographic history of European populations *Arabidopsis thaliana*,” *PLoS Genetics*, 4, e1000075.
- Grelaud, A., Robert, C.P., Marin, J.-M., Rodolphe, F., and Taly, J.-F. (2009), “ABC likelihood-free methods for model choice in Gibbs random fields,” *Bayesian Analysis*, 4, 317–336.
- Jabot, F., and Chave, J. (2009), “Inferring the parameters of the neutral theory of biodiversity using phylogenetic information and implications for tropical forests,” *Ecology Letters*, 12, 239–248.
- Joyce, P., Genz, A., and Buzbas, E.O. (2012), “Efficient simulation and likelihood methods for non-neutral multi-allele models,” *Journal of Computational Biology*, 19, 650–661.
- Joyce, P., and Marjoram, P. (2008), “Approximately sufficient statistics and Bayesian computation,” *Statistical Applications in Genetics and Molecular Biology*, 7, Article 26.
- Kotz, S., Balakrishnan, N., and Johnson, N.L. (2000), *Continuous Multivariate Distributions, Second Ed.*, Wiley-Interscience, New York.
- Leuenberger, C., and Wegmann, D. (2010), “Bayesian computation and model selection without likelihoods,” *Genetics*, 184, 243–252.
- Liu, J.S. (2008), *Monte Carlo Strategies in Scientific Computing*, Springer, New York.
- Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S. (2003), “Markov chain Monte Carlo without likelihoods,” *Proceedings of the National Academy of Sciences*, 100, 15324–15328.

- Marjoram, P., and Tavaré, S. (2006), “Modern computational approaches for analysing molecular genetic variation data,” *Nature Reviews Genetics*, 7, 759–770.
- Marin, J.-M., Pudlo, P., Robert, C.P., and Ryder, R. (2011), “Approximate Bayesian computational methods,” *Statistics and Computing*, 22, 1167–1180.
- Nunes, M.A., and Balding, D.J. (2010), “On optimal selection of summary statistics for approximate Bayesian computation,” *Statistical Applications in Genetics and Molecular Biology*, 34, Article 34.
- Owen, A. (1990), “Empirical likelihood ratio confidence regions,” *Annals of Statistics*, 18, 90–120.
- Plagnol, V., and Tavaré, S. (2004), “Approximate Bayesian computation and MCMC,” in *Monte Carlo and Quasi-Monte Carlo Methods*, ed. H. Niederreiter, Springer-Verlag, pp 99–114.
- Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A., and Feldman, M.W. (1999), “Population growth of human Y chromosomes: a study of Y chromosome microsatellites,” *Molecular Biology and Evolution*, 16, 1791–1798.
- Ratmann, O., Andrieu, C., Wiuf, C., and Richardson, S. (2009), “Model criticism based on likelihood-free inference, with an application to protein network evolution,” *Proceedings of the National Academy of Sciences*, 106, 10576–10581.
- Robert, C.P., Cornuet, J.-M., Marin, J.-M., and Pillai, N.S. (2011), “Lack of confidence in approximate Bayesian computation model choice,” *Proceedings of the National Academy of Sciences*, 108, 15112–15117.
- Robert, C.P., G. Casella, (2004). *Monte Carlo Statistical Methods, 2nd Edition*. Springer, New York.
- Rubin, D.B. (1981), “Bayesian Bootstrap,” *Annals of Statistics*, 9, 130–134.
- Rubin, D.B. (1984), “Bayesianly justifiable and relevant frequency calculations for the applied statistician,” *Annals of Statistics*, 12, 1151–1172.

- Siegmund, K.D., Marjoram, P., and Shibata, D. (2008), “Modeling DNA methylation in a population of cancer cells,” *Statistical Applications in Genetics and Molecular Biology*, 7, Article 18.
- Sisson, S.A., Fan, Y., and Tanaka, M.M. (2007), “Sequential Monte Carlo without likelihoods,” *Proceedings of the National Academy of Sciences*, 104, 1760–1765.
- Sisson, S.A., Fan, Y., and Tanaka, M.M. (2009), “Correction for Sisson et al., Sequential Monte Carlo without likelihoods,” *Proceedings of the National Academy of Sciences*, 106, 16889–16889.
- Sisson, S.A., Peters, G.W., Briers, M., and Fan, Y. (2010), “A note on the target distribution ambiguity of likelihood-free samplers,” *arXiv*, 1005.5201.
- Tang, H., Peng, J., Wang, P., and Risch, N.J. (2005), “Estimation of individual admixture: Analytical and study design considerations,” *Genetic Epidemiology*, 28, 289–301.
- Tavaré, S., Balding, D.J., Griffiths, R.C., and Donnelly, P. (1997), “Inferring coalescence times from DNA sequence data,” *Genetics*, 145, 505–518.
- Tavaré, S., Marshall, C.R., Will, O., Soligo, C., and Martin, R.D. (2002), “Using the fossil record to estimate the age of the last common ancestor of extant primates,” *Nature*, 416, 726–729.
- Tavaré, S. (2005), “Ancestral inference for branching processes,” in *Branching Processes in Biology: Variation, Growth, Extinction*, eds. P. Haccou, P. Jagers, and V. Vatutin, Cambridge University Press, Cambridge, pp. 208–217.
- Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M.P.H. (2009), “Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems,” *Journal of the Royal Society, Interface*, 6, 187–202.
- Verdu, P., Austerlitz, F., Estoup, A., Vitalis, R., Georges, M., Théry, S., Froment, A., Le Bomin, S., Gessain, A., Hombert, J.-M., Van der Veen, L., Quintana-Murci, L., Bahuchet, S., and Heyer, E. (2009), “Origins and

- genetic diversity of Pygmy hunter-gatherers from western Central Africa,” *Current Biology*, 19, 312–318.
- Verdu, P., and Rosenberg, N.A. (2011), “A general mechanistic model for admixture histories of hybrid populations,” *Genetics*, 189, 1413–1426.
- Walker, D. M., Allingham, D., Lee, H.W.J., and Small, M. (2010), “Parameter inference in small world network disease models with approximate Bayesian computational methods,” *Physica A, Statistical Mechanics and its Applications*, 389, 540–548.
- Weiss, G., and Von Haeseler, A. (1998), “Inference of population history using a likelihood approach,” *Genetics*, 149, 1539–1546.
- Wegmann, D., Leuenberger, C., and Excoffier, L. (2009), “Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood,” *Genetics*, 182, 1207–1218.
- Wilkinson, R.D. (2008), “Approximate Bayesian computation (ABC) gives exact results under the assumption of model error,” *arXiv*, 0811.3355v1.
- Wilkinson, R.D., Steiper, M., Soligo, C., Martin, R., Yang, Z., and Tavaré, S. (2010), “Dating primate divergences through an integrated analysis of palaeontological and molecular data,” *Systematic Biology*, 60, 16–31.
- Wright, S. (1949), “Adaptation and selection,” in *Genetics, Paleontology, and Evolution*, eds. G.L. Jepsen, G.G. Simpson, and E. Mayr., Princeton University Press, Princeton, NJ.

Appendix 1

We let $k \leq n$ be the number of distinct values $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k$ in the data set $\tilde{\mathbf{x}}$, and denote the number of observed \tilde{x}_i by \tilde{n}_i , where $n = \sum_{i=1}^k \tilde{n}_i$. Then the prior distribution for the probabilities of an AABC replicate data set based on the ABC simulated data set $\tilde{\mathbf{x}}$ is the Dirichlet distribution $\pi(\boldsymbol{\phi}) = [\Gamma(\sum_{i=1}^k \tilde{n}_i) / \prod_{i=1}^k \Gamma(\tilde{n}_i)] \prod_{i=1}^k \phi^{\tilde{n}_i - 1}$ with parameters $\tilde{n}_1, \tilde{n}_2, \dots, \tilde{n}_k$. The special case of the prior proportional to 1 described in the text is obtained with $k = n$, when all observations in $\tilde{\mathbf{x}}$ are distinct ($\tilde{n}_1 = \tilde{n}_2 = \dots = \tilde{n}_n = 1$). Our goal is to show that $\lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \pi_\epsilon(\theta | \mathbf{x}_o, Q_\theta) = \lim_{n \rightarrow \infty} \pi_\epsilon(\theta | \mathbf{x}_o, P_\theta)$.

Recalling equation 4,

$$\lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \pi_\epsilon(\theta | \mathbf{x}_o, Q_\theta) = \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{1}{C_{Q_\theta}} \int_{\mathcal{X}} \mathbf{I}_{\{\|\mathbf{s} - \mathbf{s}_o\| < \epsilon\}} \left[\int_{\Phi} q(\mathbf{x} | \phi, \tilde{\mathbf{x}}) \pi(\phi) d\phi \mathbf{I}_{\{\theta, \tilde{\theta}\}} \right] \pi(\theta) d\mathbf{x}. \quad (6)$$

The integral in the brackets is the expectation of $q(\mathbf{x} | \phi, \tilde{\mathbf{x}})$, with respect to the prior $\pi(\phi)$. We let $C = \binom{n}{n_1 \ n_2 \ \dots \ n_k}$, and using the definition of $q(\mathbf{x} | \phi, \tilde{\mathbf{x}}) = C \prod_{j=1}^n \prod_{i=1}^k \phi_i^{\mathbf{I}_{\{x_j = \tilde{x}_i\}}}$ in section 3.1, and $\pi(\phi) = [\Gamma(\sum_{i=1}^k \tilde{n}_i) / \prod_{i=1}^k \Gamma(\tilde{n}_i)] \prod_{i=1}^k \phi_i^{\tilde{n}_i - 1}$ we get

$$\int_{\Phi} q(\mathbf{x} | \phi, \tilde{\mathbf{x}}) \pi(\phi) d\phi = C \frac{\Gamma(\sum_{i=1}^k \tilde{n}_i)}{\prod_{i=1}^k \Gamma(\tilde{n}_i)} \prod_{j=1}^n \int_{\Phi} \left(\prod_{i=1}^k \phi_i^{\mathbf{I}_{\{x_j = \tilde{x}_i\}}} \right) \left(\prod_{i=1}^k \phi_i^{\tilde{n}_i - 1} \right) d\phi.$$

Here, we have exchanged the order of the product over j with the integral since the expectation of the product of n IID observations in sample \mathbf{x} is equal to the the product of the expectations of observations x_j . We label the realized value of the j th data point x_j by (j) such that $\prod_{i=1}^k \phi_i^{\mathbf{I}_{\{x_j = \tilde{x}_i\}}} = \phi_{(j)}$, and write

$$\int_{\Phi} q(\mathbf{x} | \phi, \tilde{\mathbf{x}}) \pi(\phi) d\phi = C \frac{\Gamma(\sum_{i=1}^k \tilde{n}_i)}{\prod_{i=1}^k \Gamma(\tilde{n}_i)} \prod_{j=1}^n \int_{\Phi} \left(\prod_{\substack{i=1 \\ i \neq (j)}}^k \phi_i^{\tilde{n}_i - 1} \right) \phi_{(j)}^{\tilde{n}_{(j)}} d\phi. \quad (7)$$

Using $\int_{\Phi} \frac{\Gamma[(\sum_{i=1, i \neq (j)}^k \tilde{n}_i) + \tilde{n}_{(j)} + 1]}{[\prod_{i=1, i \neq (j)}^k \Gamma(\tilde{n}_i)] \Gamma(\tilde{n}_{(j)} + 1)} \left(\prod_{i=1, i \neq (j)}^k \phi_i^{\tilde{n}_i - 1} \right) \phi_{(j)}^{\tilde{n}_{(j)}} d\phi = 1$ (p. 487, Kotz *et al.* (2000)), we substitute the integral in equation (7) with the ratio of the gamma functions to get

$$\begin{aligned} \int_{\Phi} q(\mathbf{x} | \phi, \tilde{\mathbf{x}}) \pi(\phi) d\phi &= C \frac{\Gamma(\sum_{i=1}^k \tilde{n}_i)}{\prod_{i=1}^k \Gamma(\tilde{n}_i)} \prod_{j=1}^n \frac{[\prod_{i=1, i \neq (j)}^k \Gamma(\tilde{n}_i)] \Gamma(\tilde{n}_{(j)} + 1)}{\Gamma[(\sum_{i=1, i \neq (j)}^k \tilde{n}_i) + \tilde{n}_{(j)} + 1]} \\ &= C \prod_{j=1}^n \frac{\Gamma(n)}{\Gamma(\tilde{n}_{(j)})} \frac{\Gamma(\tilde{n}_{(j)} + 1)}{\Gamma(n + 1)} = C \prod_{j=1}^n \left(\frac{\tilde{n}_{(j)}}{n} \right). \end{aligned}$$

Substituting $C \prod_{j=1}^n \left(\frac{\tilde{n}_{(j)}}{n} \right)$ for the integral in brackets in equation (6), we

have

$$\begin{aligned}
\lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \pi_\epsilon(\theta | \mathbf{x}_o, Q_\theta) &= \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{1}{C_{Q_\theta}} \int_{\mathcal{X}} \mathbf{I}_{\{\|s - s_o\| < \epsilon\}} C \prod_{j=1}^n \left(\frac{\tilde{n}_{(j)}}{n} \right) \mathbf{I}_{\{\theta, \tilde{\theta}\}} \pi(\theta) d\mathbf{x} \\
&= \frac{\lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \int_{\mathcal{X}} \mathbf{I}_{\{\|s - s_o\| < \epsilon\}} C \prod_{j=1}^n \left(\frac{\tilde{n}_{(j)}}{n} \right) \mathbf{I}_{\{\theta, \tilde{\theta}\}} \pi(\theta) d\mathbf{x}}{\lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} C_{Q_\theta}}.
\end{aligned} \tag{8}$$

We apply the dominated convergence theorem to exchange the limits in n and the integrals in the numerator and denominator of equation (8). The assumptions of the theorem are satisfied as follows: 1) The integrand in equation (8) is bounded: The indicator functions are bounded by 1, the ratios $(\tilde{n}_{(j)}/n)$, where $n_{(j)} \leq n$ are bounded by 1, and the prior $\pi(\theta)$ is bounded by assumption. 2) $\lim_{n \rightarrow \infty} (\tilde{n}_{(j)}/n)$ converges pointwise to the probability of $x_{(j)}$ under $\tilde{\theta}$ and the model $P_{\tilde{\theta}}$, given by $p(x_{(j)}|\tilde{\theta})$, by the frequency interpretation of probability. Exchanging the limits in n and the integrals, and using $\lim_{n \rightarrow \infty} (\tilde{n}_{(j)}/n) = p(x_{(j)}|\tilde{\theta})$,

$$\begin{aligned}
\lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \pi_\epsilon(\theta | \mathbf{x}_o, Q_\theta) &= \frac{\lim_{m \rightarrow \infty} \int_{\mathcal{X}} \mathbf{I}_{\{\|s - s_o\| < \epsilon\}} \prod_{j=1}^k \left[p(x_{(j)}|\tilde{\theta}) \right]^{n_{(j)}} \mathbf{I}_{\{\theta, \tilde{\theta}\}} \pi(\theta) d\mathbf{x}}{\lim_{m \rightarrow \infty} C_{P_{\tilde{\theta}}}} \\
&= \frac{\lim_{m \rightarrow \infty} \int_{\mathcal{X}} \mathbf{I}_{\{\|s - s_o\| < \epsilon\}} p(\mathbf{x}|\tilde{\theta}) \mathbf{I}_{\{\theta, \tilde{\theta}\}} \pi(\theta) d\mathbf{x}}{\lim_{m \rightarrow \infty} C_{P_{\tilde{\theta}}}},
\end{aligned} \tag{9}$$

where (9) follows by the definition of the joint distribution $p(\mathbf{x}|\tilde{\theta}) = \prod_{j=1}^k \left[p(x_{(j)}|\tilde{\theta}) \right]^{n_{(j)}}$.

We now apply the dominated convergence theorem a second time to exchange the limits in m and the integrals on \mathcal{X} . Again, the assumptions of the dominated convergence theorem are satisfied since the integrand in (9) is a sequence in m of bounded functions, and as $m \rightarrow \infty$, $\tilde{\theta} \rightarrow \theta$, and $p(\mathbf{x}|\tilde{\theta}) \rightarrow p(\mathbf{x}|\theta)$. We get

$$\lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \pi_\epsilon(\theta | \mathbf{x}_o, Q_\theta) = \frac{1}{C_{P_\theta}} \int_{\mathcal{X}} \mathbf{I}_{\{\|s - s_o\| < \epsilon\}} p(\mathbf{x}|\theta) \pi(\theta) d\mathbf{x} = \lim_{n \rightarrow \infty} \pi_\epsilon(\theta | \mathbf{x}_o, P_\theta)$$

which shows that AABC posterior converges to the ABC posterior as the sample size n and the simulated number of data sets m increase.